Implementation considerations of an Expert System to assess Stream Water Quality management

Cabanillas, D. and Willmott, S.

Research Report LSI-04-32-R

Departament de Llenguatges i Sistemes Informàtics

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Implementation considerations of an Expert System to assess Stream Water Quality management

David Cabanillas [a] and Steven Willmott[a]

[a]Universitat Politècnica de Catalunya,
Edifici C5-Campus Nord, Jordi Girona 1-3,E-08034 Barcelona
Catalonia, Spain

Abstract:

The main goal of STREAMES (STream REAch Management, an Expert System) prototype is to assist to the water managers to evaluate human altered streams, via point and non–point sources and magnitudes of nutrient, deciding the best strategy in front of them.

This technical report explains how the system is composed, interrelated, treating in depth about architectural and technological aspects. Our system is an environmental system. It is a system concerning on environmental domain, concretely on rivers ecosystem.

Keywords: Expert Systems; Environmental Decision Support System; stream nutrient loads.

1 INTRODUCTION

This paper describes the implementation of the The STREAMES European project[1], in the frame of European Commission's Fifth Framework Programme (1998-2002) and Water Framework Directive (2000). It aims to analyze nutrient cycles in a particular, human-altered environment: the river ecosystem, with special emphasis on the Mediterranean region. The main objective is to assist to the water managers, helping in the decision taking. The objective of the Expert System (ES) [8] is to capture knowledge from both water managers and environmental scientists regarding nutrient effects in streams and combine their knowledge into a user–friendly decision support system [6]. Although the project refers to develope an ES the fact is that we have developed an EDSS because we have produced different related tools to help the water manager decisions,[2] [6]. The process to decide if a specific river achieve a good state involved a lot of variables such as hydrology, geomorphology, types of vegetation, soil type and land use and so forth all effect on nutrient and can change prognoses for nutrient effects in each river area. Water managers and environmental scientists often have conflicting points of view on a particular problem, furthermore rule writing is made complex by the constraints on what and how can be expressed the Expert System (ES) implementation.

The paper discusses both architectural and technology concerns, the interplay between interfaces for non–specialists and a rule based architecture [3] as well as the illicitation of knowledge that was necessary to devise the rules.

The organization of this paper is as follows: In the next section, we describe which are the important points of our project. In section 3 – 4 we draw attention to the design and technologies involved in the deployed and aspects implementation. We conclude the paper with a brief conclusion and future work and research way.

2 BACKGROUND

The water managers should evaluate a large quantity of information to determine which is the state of a river and which are the actions to realize in front of

[1]EVK1-CT-2000-00081, [http://www.streames.org]

[2]EDSS is a system under the control of one or more decision makers that assists in the activity of decision making by providing an organized set of tools intended to impart structure to portions of the decision–making situation and to improve the ultimate effectiveness of the decision outcome

[3]http://www.aaai.org/AITopics/html/expert.html

different problems[4].

In order to facilitate the tasks to the water manager we have developed a system that collect the knowledge of different sources extracted from literature and human experts as well as the specific knowledge than can only be acquired by processing empirical data directly collected from each study site. In any case the ES will offer possible options and it will be task of water managers to choose the more suitable to each case.

The current implementation is Java based (rule engine) with a Visual Basic User front–end and the captures around 250 rules for several scenarios types (making it a non–trivial Expert System). While the current prototype only covers the reach level the implementation of the system is already plenty of significant complexity and provides an instructive use case on development of a Artificial Intelligence solution in conjunction with domain experts.

The innovative character of the ES lies in its ability to estimate and evaluate uncertainties in the prediction of the stream water quality in support of management decisions. Another important issue is the integration of a Geographic Information System (GIS) [5] linked to a nutrient export model MOdelling Nutrient Emission in RIver Systems MONERIS [2] into the development of the ES, to address spatial information at the appropriate stream management scale.

3 DESIGN

The idea is to design an EDSS that follows the next features:

- modular: The EDSS is formed by some pieces such as: ES, MONERIS model, GIS
- reusable: The ES is a open shell, so may be used in other environments. Only will be necessary to re–write the rules (namely the knowledge)
- expandable: Streames have been a prototype, probably in a next future it will be necessary to include more functionalities

The more important aspects in the design choices have been to select the knowledge base systems [7] more adequate, the next list describe the three possible paradigms:

- Rule–based heuristic (expert) knowledge encoded in rules.
- Model–based reasoning is based on a model of the system.
- Case-based knowledge is provided by many examples of solutions to previous cases.

We have chose the first one, the reasons are because:

- Rule–based systems are a relatively simple model that can be adapted to any number of problems
- In some cases we don't dispose solutions to previous cases
- For the scientifics is suitable to write his / her knowledge to rules

In the next sections, we describe the most important facet of the project.

3.1 Obtain / Representation / Organization of the Knowledge

Knowledge acquisition ([1], [5]) is the stone corner of whatever rule based engine. Without the expert's knowledge is not possible to write down rules that describe the behavior of the system to treat.

In our case we have steering with a great number of information sources. First at all we have two kinds of institutions – research and management– one of them having a particular problem view / knowledge and achieve ours goals. Furthermore, the participants come from 17 institutions in 8 countries and have been selected eleven different sites. All this great amount of heterogenous knowledge has been obtained by means of forms, interviews, meetings with the water managers and environmental scientists.

The syntax to describe the rules is naif but enough powerful to describe our knowledge. We dispose of a different decision trees (each one for each one problem). Decision Trees has the following advantages:

- Outputs the hypothesis in the form of a tree

[4] Following the 2000/60/EEC Water Framework Directive
[5] www.gis.com

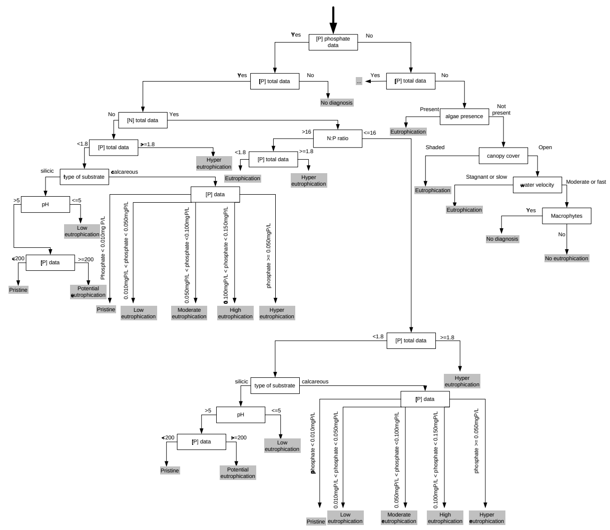

Figure 1: Decision Tree

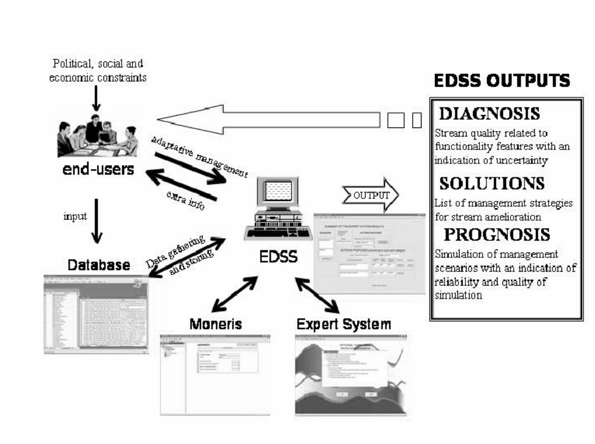

Figure 2: Architecture design

- Can be though of as rules, or disjunction of conjunction of constraints
- Very popular in data-mining because it is easy to understand algorithm and output

With this bunch of IF–THEN rules and the information introduced the system is prepared to run. The knowledge organization is to make groups of rules / facts related by problems. It means that Eutrophication problem[6] has related a set of rules and facts stored in a unique table. With this approach is easier to know which rules are triggers and to modify the rules in the future. Because the end–user only needs to modify one table and each one only works over a problem. For more details is advisable to review [5].

3.2 Architecture design

We have divided the architecture 2 in three parts. First, the Exper System (ES) which is considered the heart of our system. In a second part the interface that it's showed to the end–user. And finally the relation between the input data user and the results obtained by the ES.

Expert System

The ES architecture is composed by:

[6] Eutrophication: gradual increase in the concentration of nutrients in an aging aquatic ecosystem.

- Knowledge base that contains facts and rules, represented using a formalism (semantic nets, frames,...)
- Inference engine
- Applies rules to facts and produces new knowledge

For more details particular emphasis on the implementation of the rule system governing the central part of the reasoning. In [4] for BESAI 2004 [7] in process of revision.

Inference engine:

- Interprets rules against facts using a strategy
- Strategy
- detection: finds "firing" rules
- selection: solves conficts
- application: inference (forward chaining in our case)

Selection: From the rules identified during the detection stage, find the best one. Solve conflicts, with various criteria implemented such as:

- use the most popular rule
- most specific / general

[7] http://www.lsi.upc.es/ webia/besai/besai2004.html

- the most informative (high number of new facts)

Application:

- Execute the selected rule
- Add new facts into the rule base
- Propagate rules
- Check for termination (if all rules have been applied, or a conclusion has been reached)

An Expert System (ES) is a computer program that performs difficult and specialized tasks at the level of human expert. The structure of the ES presents two main independent modules: the Knowledge Base (KB) and the Inference Engine (IE). The Figure 2 shows out the different parts involved in the project. In paper [3] for iEMSs 2004 [8] in revision, where we summarize the decision–support knowledge components which have been identified in previous work and, based on these, present an implementation of a prototype of an environmental decision–support system.

Interface

The interface 5 is the input data door and we have treated that this task comes easy and offers a lot of help in order to assure to the user a confidence in the system.

Interconnection between ES and user

The interconnection between the Expert System and the final user are showed in two ways. In Figure 3 the relation from the input data by the end–user and Figure 4 the opposite way from the inference engine to end–user.

3.3 Technologies Choices

So far, we have been describing the aspect more related with the concept level in technological aspects. Once you have the design, the next step, commonly, is to implement it. During this process it is necessary to look after to chose the most adequate technologies that in each case are necessary in another way we may have a negative impact on the progress of the project.

[8]http://www.iemss.org/iemss2004/

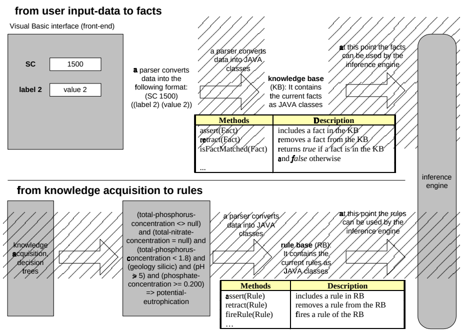

Figure 3: From input data to end–user

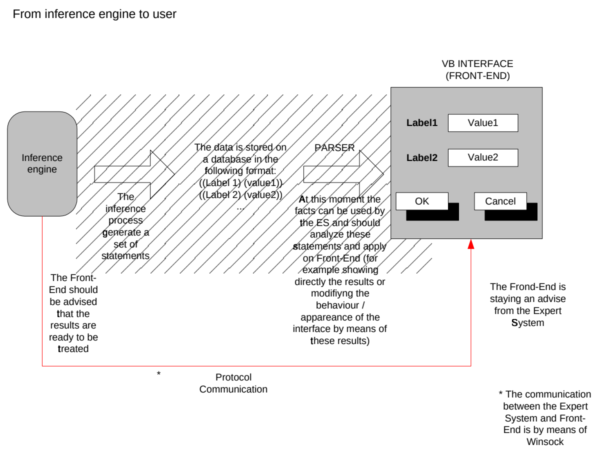

Figure 4: From inference engine to end–user

About our proposal solution. It is possible to devise two important groups of Expert Systems:

- Commercial shell (like such Jess[9] or G2[10]) and
- Open source software (like such CLIPS[11] or Mandarax[12])

Due to the features imposed in this project we didn't use a commercial shell. We have obtained knowledge about which is the behavior / structure of an expert system and we have deployed a new one following traditional theories. The more important advantage of this approach is that we know how it works and in a future it can be improved / expanded.

[9]herzberg.ca.sandia.gov/jess/
[10]www.gensym.com
[11]www.ghg.net/clips/CLIPS.html
[12]www.mandarax.org/

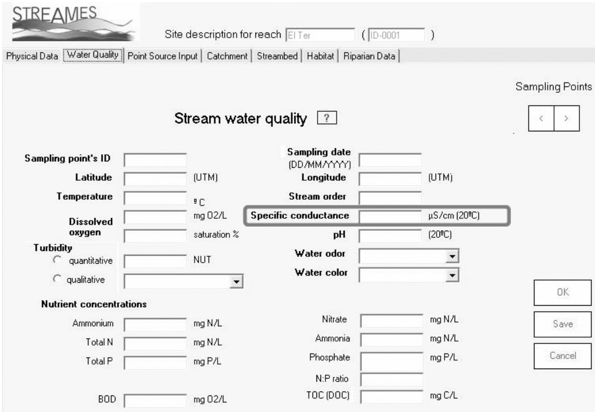

Figure 5: 1-conductance-interface

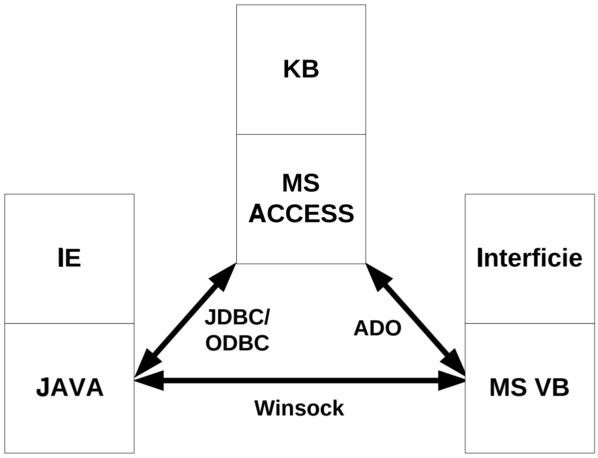

Figure 6: Technologies choices

About the programming languages: We have considered two candidates for the programming language to develop our ES, which were Visual Basic (VB) and Java. We have used VB to deploy the user interface (like front–end 5), while we have used Java (like back–end) which is platform independent, to deploy the inference engine. The idea is to deploy a reusable inference engine and to bind it with the specific interface and knowledge base for each one case.

About the Databases: A database is essentially a repository of data, which allows information to be stored and retrieved quickly. Our options had been Microsoft Access and the main reason is due to its ability to integrate with other Microsoft technologies and its simplicity to set up. However, we would like to consider another possibilities such as Microsoft SQL Server or MySQL.

About the Communication: Due that we have selected two different languages it is necessary to communicate them. We have considered WinSock, which provides a two-way communication between applications either on the same computer or two different computers, to achieve the interoperation of systems between them.

In the diagram 6 we devise the different parts and how are related.

4 IMPLEMENTATION

Implementation and test:

- Implement the knowledge base
- Implement the inference engine
- Test on a set of known cases

5 TESTING AND RESULTS

The validation involved two different aspects: a validation of the technical aspects and a validation of the outputs obtained by the EDSS. The first one have been realized during the stage of developing and the second we will compare the outputs by EDSS with a well known examples. Furthermore, the last Work Package is a dissemination of the systems and we expect to obtain more information during this phase about which are the feelings of the end–user in front the system.

6 CONCLUSIONS AND FUTURE WORK

This paper describes the more technicians aspects during the development of STREAMES project.

The next list describes possible future ways:

- Relation between other technologies Moneris / GIS to improve the EDSS
- Introducing more power in the rules' syntax (fuzzy sets, operators)
- Automatic rule generation and validation
- Compare with other knowledge base systems such as CBR systems, MBR systems.

REFERENCES

[1] J.-L. Aguirre, O. Barron, R. Brena, and M.-N. Garcia. Using a high level knowledge representation for expert systems knowledge acquisition and prototyping. In *Proceedings of the 1993 ACM/SIGAPP symposium on Applied computing*, pages 490–497. ACM Press, 1993.

[2] H. Behrendt. The model system moneris. http://www.iia-cnr.unical.it/EUROCAT/Pag.CNR/publications9.pdf, 2003.

[3] D. Cabanillas, J. Comas, E. Llorens, M. Poch, and L. C. S. Willmott. Implementation of the streames environmental decision–support system.

[4] L. Ceccaroni, D. Cabanillas, and J. Comas. Symptom discovery and problem diagnosis in the streames environmental decisionsupport system, 2004.

[5] J. Comas, L. E., E. Mart, M. A. Puig, J. L. Riera, F. Sabater, and M. Poch. Knowledge acquisition in the streames project: the key process in the environmental decision support system development. *AI Communications*, 16:253–265, 2003.

[6] G. M. Marakas. *Decision support systems in the twenty-first century*. Prentice-Hall, Inc., 1999.

[7] C. Mérida-Campos and E. Rollón. Cbr and mbr techniques: review for an application in the emergencies domain. Technical Report LSI-03-36-R, Software Departament. Technical University of Catalonia. Barcelona Spain, 2003.

[8] E. Turban and P. R. Watkins. *Applied Expert Systems*. Elsevier Science Inc., 1988.

Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya

Research Reports - 2004

- LSI-04-1-R : *Automatic Generation of Polynomial Loop Invariants: Algebraic Foundations*, Rodríguez, E. and Kapur, D.
- LSI-04-2-R : *Comparison of Methods to Predict Ozone Concentration* , Orozco, J.
- LSI-04-3-R : *Towards the definition of a taxonomy for the cots product´s market* , Ayala, Claudia P.
- LSI-04-4-R : *Modelling Coalition Formation over Time for Iterative Coalition Games*, Mérida-Campos, C. and Willmott, S.
- LSI-04-5-R : *Illegal Agents? Creating Wholly Independent Autonomous Entities in Online Worlds*, Willmott, S.
- LSI-04-6-R : *An Analysis Pattern for Electronic Marketplaces*, Queralt, A. and Teniente, E.
- LSI-04-7-R : *Exploring Dopamine-Mediated Reward Processing through the Analysis of EEG-Measured Gamma-Band Brain Oscillations*, Vellido, A. and El-Deredy, W.
- LSI-04-8-R : *Studying Embedded Human EEG Dynamics Using Generative Topographic Mapping*, Vellido, A. and El-Deredy, W. and Lisboa, P.J.G.
- LSI-04-9-R : *Similarity and Dissimilarity Concepts in Machine Learning*, Orozco, J.
- LSI-04-10-R : *A Framework for the Definition of Metrics for Actor-Dependency Models*, Quer, C. and Grau, G. and Franch, X.
- LSI-04-11-R : *QM: A Tool for Building Software Quality Models*, Carvallo, J.P. and Franch, X. and Grau, G. and Quer, C.
- LSI-04-12-R : *COSTUME: A Method for Building Quality Models for Composite COTS-based Software Systems*, Carvallo, J.P. and Franch, X. and Grau, G. and Quer, C.
- LSI-04-13-R : *Enabling Collaboration in Virtual Reality Navigators*, Theoktisto, V. and Fairén, M. and Navazo, I.
- LSI-04-14-R : *DesCOTS: A Software System for Selecting COTS Components*, Carvallo, J.P. and Franch, X. and Grau, G. and Quer, C.
- LSI-04-15-R : *Evaluation and symmetrisation of alignments obtained with the Giza++ software*, Lambert, P. and Castell, N.
- LSI-04-16-R : *A note on the use of topology extensions for provoking instability in communication networks*, Blesa, M.J.
- LSI-04-17-R : *An ISO/IEC-compliant Quality Model for ER Diagrams*, Costal, D. and Franch, X.
- LSI-04-18-R : *A Case Study on Pruning General Ontologies for the Development of Conceptual Schemas* , Conesa, J.
- LSI-04-19-R : *Adding Efficient and Reliable Access Paths to the JCF*, Marco, J. and Franch, X.

- LSI-04-20-R : *Exploiting Simple Corporate Memory in Iterative Coalition Games*, Mérida-Campos, C. and Willmott, S.
- LSI-04-21-R : *On the Semantics of Operation Contracts in Conceptual Modeling* , Queralt, A. and Teniente, E.
- LSI-04-22-R : *Complexity issues on bounded restrictive H-coloring*, Díaz, J. and Serna, M. and Thilikos, D.M.
- LSI-04-23-R : *Chromatic number in random scaled sector graphs*, Díaz, J. and Sanwalani, V. and Serna, M. and Spirakis, P.
- LSI-04-24-R : *Bounds on the bisection width for random d-regular graphs*, Díaz, J. and Serna, M. and Wormald, N.C.
- LSI-04-25-R : *Open Source environment to define constraints in route planning for GIS-T*, Pérez, L. and Silveira, A. da M.
- LSI-04-26-R : *A basic repository of operations for the refinement of general ontologies*, de Palol, X.
- LSI-04-27-R : *Tetrahedral mesh subdivision based on underlying volume data*, Rodríguez, L. and Navazo, I. and Vinacua, A.
- LSI-04-28-R : *The Price of Connectedness in Expansions*, Fomin, F.V. and Fraigniaud, P. and Thilikos, D.M.
- LSI-04-29-R : *Smaller kernels for hitting set problems of constant arity*, Nishimura, N. and Ragde, P. and Thilikos, D.M.
- LSI-04-30-R : *Searching Spatial Sense in the Ontological World: Discovering Spatial Objects*, Morocho, V and Pérez, L. and Saltor, F.
- LSI-04-32-R : *Implementation considerations of an Expert System to assess Stream Water Quality management*, Cabanillas, D. and Willmott, S.

Hardcopies of reports can be ordered from:

Núria Sanchez
Departament de Llenguatges i Sistemes Informàtics
Universitat Politècnica de Catalunya
Campus Nord, Mòdul C6
Jordi Girona Salgado, 1-3
03034 Barcelona,Spain
nurias@lsi.upc.es

See also the Departament WWW pages, http://www.lsi.upc.es/